Chapter 2 Big Data Architecture Components

ABSTRACT

The previous chapter overviewed big data including its types, sources, analytic techniques, and applications. This chapter briefly discusses the architecture components dealing with the huge volume of data. The complexity of big data types defines a logical architecture with layers and high-level components to obtain a big data solution that includes data sources with the relation to atomic patterns. The dimensions of the approach include volume, variety, velocity, veracity, and governance. The diverse layers of the architecture are big data sources, data massaging and store layer, analysis layer, and consumption layer. Big data sources are data collected from various sources to perform analytics by data scientists. Data can be from internal and external sources. Internal sources comprise transactional data, device sensors, business documents, internal files, etc. External sources can be from social network profiles, geographical data, data stores, etc. Data massage is the process of extracting data by preprocessing like removal of missing values, dimensionality reduction, and noise removal to attain a useful format to be stored. Analysis layer is to provide insight with preferred analytics techniques and tools. The analytics methods, issues to be considered, requirements, and tools are widely mentioned. Consumption layer being the result of business insight can be outsourced to sources like retail marketing, public sector, financial body, and media. Finally, a case study of architectural drivers is applied on a retail industry application and its challenges and usecases are discussed.

DOI: 10.4018/978-1-5225-3790-8.ch002

INTRODUCTION

Long time ago the research and development on infrastructure services included High speed networking for data intensive applications with collaborative systems like grids with dynamic provisioned security (Williams, 2016). Currently this is provided by software defined collaborative systems that is intercloud and Big Data technologies. There requires an architecture to be defined since the data exceeded the processing capacity of conventional systems. Hence, an alternative architecture is to be needed for processing to gain value from this data. (Sathic, 2013). The architecture framework stated to be cost-effective with innovative forms of processing for better insight to perform decision making is done.

BACKGROUND

Big Data can be stored, retrieved, processed and analysed in various ways. This includes many dimensions and requires a high computation model with security and governance. The choice of such an architecture pattern is a challenging task across huge factors. The complexity of Big Data types defines a logical architecture with layers and high level components to obtain a Big Data solution. The logical architecture includes a set of data sources and is relation with atomic patterns by focusing on each aspect for a Big Data solution.

With the beginning of Big Data technologies, organizations started querying, "What kind of insight are possible for business, governance if Big Data technologies comes into existence?" A structured approach is defined based on the dimensions to assess the feasibility of Big Data solution. The dimensions in this approach may include:

- Volume of the data
- Variety of data sources, types, and formats
- Velocity at which the data is generated, i.e. the speed
- Veracity which is uncertainty or trustworthiness of the data
- Business value from analyzing the data
- Governance for the new sources of data and its usage

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: <u>www.igi-</u> <u>global.com/chapter/big-data-architecture-</u> components/216597

Related Content

Reserve Capacity of Mixed Urban Road Networks, Network Configuration and Signal Settings

Masoomeh Divsalar, Reza Hassanzadeh, Iraj Mahdaviand Nezam Mahdavi-Amiri (2018). *Intelligent Transportation and Planning: Breakthroughs in Research and Practice (pp. 883-906).*

www.irma-international.org/chapter/reserve-capacity-of-mixed-urban-road-networks-networkconfiguration-and-signal-settings/197166

Ensuring Food Supply and Security During Localised Lockdowns: An Information and Technology-Based Approach

Iman Hussain, Chloë Allen-Ede, Lukas Jaksand Herbert Daly (2021). *Data Science Advancements in Pandemic and Outbreak Management (pp. 160-193).* www.irma-international.org/chapter/ensuring-food-supply-and-security-during-localised-lockdowns/275097

Optimized Crossover JumpX in Genetic Algorithm for General Routing Problems: A Crossover Survey and Enhancement

Hicham El Hassani, Said Benkachchaand Jamal Benhra (2018). Advancements in Applied Metaheuristic Computing (pp. 205-230).

www.irma-international.org/chapter/optimized-crossover-jumpx-in-genetic-algorithm-for-generalrouting-problems/192007

Commercial Banks' Digital Paradigm and Customers Responses in the UAE

Muhammad Jumaa (2020). International Journal of Data Analytics (pp. 68-79). www.irma-international.org/article/commercial-banks-digital-paradigm-and-customersresponses-in-the-uae/244170

Unstructured Healthcare Data Archiving and Retrieval Using Hadoop and Drill

Hang Yue (2018). *International Journal of Big Data and Analytics in Healthcare (pp. 28-44).*

www.irma-international.org/article/unstructured-healthcare-data-archiving-and-retrieval-using-hadoop-and-drill/223165